Lecture 6
Approximation Techniques

1 The Residual

Consider an equation of the following generic form:

$$A\ f(\overrightarrow{x}) = g(\overrightarrow{x})\,, \tag{1}$$

where A is a linear operator, $f(\overrightarrow{x})$ is the solution, and $g(x)$ is the driving or source function. Let us assume an approximate trial-space expansion for $f(\overrightarrow{x})$ as follows:

$$f(\overrightarrow{x}) \approx \tilde{f}(\overrightarrow{x}) = \sum_{i=1}^{N} f_i B_i(\overrightarrow{x})\,. \tag{2}$$

We further assume that the trial-space functions are chosen so that $f(\overrightarrow{x})$ naturally meets any required boundary conditions. Next we define the residual function as follows:

$$R(\overrightarrow{x}) = g(\overrightarrow{x}) - A\,\tilde{f}(\overrightarrow{x})\,. \tag{3}$$

We would like to choose the expansion coefficients $\{f_i\}_{i=1}^{N}$ so as to make $R(\overrightarrow{x})$ identically zero because then $\tilde{f}(\overrightarrow{x})$ would exactly solve Eq. (1). However, this is obviously almost never possible. The error in the solution is proportional to the size of the residual. To show this, we apply A^{-1} to Eq. (3), and slightly manipulate the resulting equation to obtain

$$f(\overrightarrow{x}) - \tilde{f}(\overrightarrow{x}) = A^{-1}R(\overrightarrow{x})\,, \tag{4}$$

where $f(\overrightarrow{x}) = A^{-1}g(\overrightarrow{x})$. Thus if we can't make the residual identically zero, we should try to make it "small" in some sense.

2 Three Basic Methods

There are several different ways to make the residual "small":

- The Least-Squares Method: choose the expansion coefficients so that the integral of the square of the residual is minimized, i.e. minimize

$$\Gamma = \sum_{\overrightarrow{x}} R^2(\overrightarrow{x})\, d\overrightarrow{x}\,. \tag{5}$$

- The Galerkin Method: choose the expansion coefficients so that the residual is orthogonal over the problem domain to a set of N linearly-independent functions $\left\{W_i(\overrightarrow{x})\right\}_{i=1}^{N}$, i.e.,

$$\int_{\overrightarrow{x}} R(\overrightarrow{x}) W_i(\overrightarrow{x})\, d\overrightarrow{x} = 0\,, \quad i = 1, N. \tag{6}$$

  The weighting functions form an N-dimensional space of functions called the weighting space. The trial space can be identical to the weighting space, i.e.,

$$B_i(\overrightarrow{x}) = W_i(\overrightarrow{x})\,, \tag{7}$$

  or it can be a different space.

- The Collocation Method: choose the expansion coefficients so that the residual is zero at N distinct points $\left\{ \overrightarrow{x}_i \right\}_{i=1}^{N}$:

$$R(\overrightarrow{x}_i) = 0\,, \quad i = 1, N. \tag{8}$$

It is clear as to why the residual is made "small" using the least-squares method and the collocation method, but it is less clear for the weighted residual method. The explanation is as follows. If the residual is orthogonal to the weighting functions, the least-squares fit to the residual using the weighting functions is identically zero. Thus the residual must be "near zero". For instance, we represent the least-squares fit as

$$\tilde{R}(\overrightarrow{x}) = \sum_{i=1}^{N} R_i W_i(\overrightarrow{x})\,. \tag{9}$$

The expansion coefficients are chosen to minimize the following functional:

$$\Gamma = \int_X [R(\overrightarrow{x}) - \sum_{i=1}^{N} R_i W_i(\overrightarrow{x})]^2 \, d\overrightarrow{x}\,. \tag{10}$$

This minimization is achieved by requiring that $\frac{\partial \Gamma}{\partial R_i} = 0,\ i = 1, N$. This results in the following matrix equation for the vector of expansion coefficients:

$$\mathcal{W}\,(\overrightarrow{R}) = \overrightarrow{\xi}\,, \tag{11a}$$

where the elements of $\mathcal{W}$ are defined by

$$w_{i,j} = \int_X W_i(\overrightarrow{x}) W_j(\overrightarrow{x})\, d\overrightarrow{x}\,, \tag{11b}$$

and

$$\overrightarrow{R} = (R_1, R_2, \cdots R_N) \,, \tag{11c}$$

$$\overrightarrow{\xi} = (\xi_1, \xi_2, \cdots \xi_N) \,, \tag{11d}$$

$$\xi_i = \int_X R(\overrightarrow{x}) W_i(\overrightarrow{x}) \, d\overrightarrow{x} \,. \tag{11e}$$

Thus the vector of expansion coefficients is given by

$$\overrightarrow{R} = \mathcal{W}^{-1} \, \overrightarrow{\xi} \,. \tag{11f}$$

If the residual is orthogonal to the weighting function, the $\overrightarrow{\xi}$ vector is identically zero, and it follows from Eq. (11f) that the expansion coefficients must also be zero.

3 Examples

Consider the following equation:

$$\frac{\partial f}{\partial x} + \sigma f = 0 \,, \tag{12}$$

which is defined over the interval, $[0, x_0]$, with the boundary condition, $f(0) = 1$. The solution to this equation is

$$f(x) = \exp(-\sigma x) \,. \tag{13}$$

The Taylor-series expansion about $x = 0$ is

$$f(x) = 1 - \tau + \frac{1}{2}\tau^2 - \frac{1}{6}\tau^3 + \frac{1}{24}\tau^4 + O(\tau^5) \,, \tag{14}$$

where $\tau = \sigma x$.

3.1 A Least-Squares Example

We next approximately solve Eq. (12) using the least-squares method in conjunction with a linear trial space that satisfies the boundary condition:

$$\tilde{f}(x) = 1 + ax\,, \tag{15}$$

where a is a constant to be determined. Substituting from Eq. (15) into Eq. (13), and forming the least-squares functional, we get:

$$\Gamma = \int_0^{x_0} \left(a + \sigma(1 + ax)\right)^2 \, dx\,. \tag{16}$$

To minimize Γ, we set $\frac{\partial \Gamma}{\partial a} = 0$:

$$\int_0^{x_0} 2\left[a + \sigma(1 + ax)\right](1 + \sigma x) \, dx = 0\,. \tag{17}$$

Manipulating Eq. (17), we get

$$\int_0^{x_0} a + \sigma + (2a\sigma + \sigma^2)x + a\sigma^2 x^2 \, dx = 0\,, \tag{18}$$

$$(a + \sigma)x_0 + (2a\sigma + \sigma^2)\frac{x_0^2}{2} + a\sigma^2\frac{x_0^3}{3} = 0\,, \tag{19}$$

$$a = -\frac{\sigma(6 + 3\sigma x_0)}{6 + 6\sigma x_0 + 2\sigma^2 x_0^2}\,. \tag{20}$$

Thus the approximate solution is

$$\tilde{f}(x) = 1 - \frac{\sigma x(6 + 3\sigma x_0)}{6 + 6\sigma x_0 + 2\sigma^2 x_0^2}\,. \tag{21}$$

Evaluating the solution at $x = x_0$ gives

$$\tilde{f}(x_0) = 1 - \frac{\sigma x_0(6 + 3\sigma x_0)}{6 + 6\sigma x_0 + 2\sigma^2 x_0^2}\,. \tag{22}$$

Expanding this solution about $x_0 = 0$ gives

$$\tilde{f}(x_0) = 1 - \tau_0 + \frac{1}{2}\tau_0^2 - \frac{1}{6}\tau_0^3 + O(\tau_0^5)\,, \tag{23}$$

where $\tau_0 = \sigma x_0$. Comparing Eq. (23) with Eq. (14), we find that the $\tilde{f}(x_0)$ is accurate through third order. This is quite good for a linear continuous approximation. On the other hand, for any cell thickness greater than approximately 2.45 mean-free-paths, the solution at x_0 is negative and therefore non-physical. Furthermore, in the limit as $x_0 \to \infty$, we find that $\tilde{f}(x_0) \to -\frac{1}{2}$, which is asymptotically incorrect and non-physical. If we integrate Eq. (12) over the interval, $[0, x_0]$, we obtain the analog of the balance equation:

$$f(x_0) - f(0) + \int_0^{x_0} \sigma f \; dx = 0\,. \tag{24}$$

If we substitute $\tilde{f}$ into Eq. (24), we find that it is not satisfied:

$$\tilde{f}(x_0) - f(0) + \int_0^{x_0} \sigma \tilde{f} \; dx = -\frac{\tau_0^3}{12 + 12\tau_0 + 4\tau_0^2}\,. \tag{25}$$

This is in keeping with the fact that least-squares methods are generally not conservative. However, also note that Eq. (25) is nonetheless met through second order in τ_0, which reflects the fact that it is a convergent method.

3.2 A Weighted Residual Example

We next approximately solve Eq. (12) using the weighted-residual method in conjunction with the linear trial space defined by Eq. (15). We will use the a weight function of unity to ensure a conservative solution. Following Eq. (6), we substitute from Eq. (15) into Eq. (12), and integrate the resulting equation over the interval, $[0, x_0]$, to obtain the equation for a:

$$\int_0^{x_0} \left[a + \sigma(1 + ax)\right] \, dx = 0 \,. \tag{26}$$

Manipulating Eq. (26), we get

$$a(x_0 + \frac{1}{2}\sigma x_0^2) + \sigma x_0 = 0 \,,$$

$$a = -\frac{2\sigma}{2 + \sigma x_0} \,. \tag{27}$$

Substituting from Eq. (27) into Eq. (15), we obtain the approximate solution:

$$\tilde{f}(x) = 1 - \frac{\sigma x}{1 + \frac{1}{2}\sigma x_0} \,. \tag{28}$$

Evaluating Eq. (28) at $x = x_0$, we get

$$\tilde{f}(x_0) = 1 - \frac{\sigma x_0}{1 + \frac{1}{2}\sigma x_0} \,. \tag{29}$$

Expanding Eq. (29) about $x_0 = 0$, we obtain

$$\tilde{f}(x_0) = 1 - \tau_0 + \frac{1}{2}\tau_0^2 - \frac{1}{4}\tau_0^3 + O(\tau_0^4)\,, \tag{30}$$

where $\tau_0 = \sigma x_0$. Comparing Eq. (29) with Eq. (14), we find that $\tilde{f}(x_0)$ is correct through second order. This is good for a linear approximation. However, for any cell thickness greater than approximately 2 mean-free-paths, the solution at x_0 is negative and therefore non-physical. Furthermore, in the limit as $x_0 \to \infty$, we find that $\tilde{f}(x_0) \to -1$, which is asymptotically incorrect and non-physical. Nonetheless if we substitute from Eq. (28) into Eq. (24), we find that the resulting equation is satisfied. This is in keeping with the fact that a weight function of unity should result in a conservative solution.

3.3 A Collocation Example

We next approximately solve Eq. (12) using the weighted-residual method in conjunction with the linear trial space defined by Eq. (15). We choose $x_0/2$ as the collocation point for reasons explained later. Following Eq. (8), we obtain the equation for the constant a:

$$a + \sigma(1 + ax_0/2) = 0\,. \tag{31}$$

Solving for a, we get

$$a = -\frac{2\sigma}{2 + \sigma x_0}\,. \tag{32}$$

Comparing Eq. (32) with Eq. (27), we find that the collocation solution is identical to the weighted-residual solution. To see why this is so, one need simply perform the integral in Eq. (26) using a one-point quadrature with the quadrature point equal to $x_0/2$, and the quadrature weight equal to $\Delta x = x_0$:

$$[a + \sigma(1 + ax_0/2)]\,\Delta x = 0\,. \tag{33}$$

Because the quadrature point is equal to $x_0/2$, the quadrature integration is exact (the midpoint rule applied to a linear integrand). Furthermore, if one divides Eq. (33) by Δx, one obtains Eq. (31). Thus in this case, collocation at $x = x_0/2$ is equivalent to an exact quadrature integration of the weighted residual equation, so the collocation method yields the weighted-residual result. This is representative of a general approach often taken with collocation. For instance, let us assume that we have collocated at quadrature points, $\{x_n\}_{n=1}^N$ with associated weights $\{w_n\}_{n=1}^N$. Then the residual satisfies

$$R(x_n) = 0\,, \quad n = 1, N. \tag{34}$$

Given a set of weight functions, $\{W_i(x)\}_{i=1}^N$, we multiply Eq. (34) by $W_i(x_n)w_n$, where i successively takes on each value from 1 to N:

$$W_i(x_n)\,R(x_n)\,w_n = 0 \quad n = 1, N, \quad i = 1, N. \tag{35}$$

Summing Eq. (35) over all n, we get

$$\sum_{n=1}^{N} W_i(x_n)\, R(x_n)\, w_n = 0 \quad i = 1, N. \tag{36}$$

If the quadrature formula is exact for the product of the residual and each weight function, it is clear from Eq. (36) that the weighted residual equations will be satisfied, and therefore that the collocation and weighted residual methods will be equivalent. If not, the collocation method should be "similar" to the weighted residual method.